



Published in final edited form as:

*Breast Cancer Res Treat.* 2011 February ; 126(1): 185–192. doi:10.1007/s10549-010-1113-7.

## Defining breast cancer prognosis based on molecular phenotypes: results from a large cohort study

**Shaheenah Dawood,**

Department of Breast Medical Oncology, Dubai Hospital, Dubai, UAE

**Rong Hu,**

Channing Laboratory, Department of Medicine, Brigham, and Women's Hospital, Harvard Medical School, Boston, MA, USA

**Michelle D. Homes,**

Channing Laboratory, Department of Medicine, Brigham, and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

**Laura C. Collins,**

Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

**Stuart J. Schnitt,**

Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

**James Connolly,**

Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

**Graham A. Colditz,** and

Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

**Rulla M. Tamimi**

Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Rulla M. Tamimi: rulla.tamimi@channing.harvard.edu

### Abstract

The objective of this study is to define the survival outcomes associated with distinct molecular phenotypes defined by immunohistochemical staining of paraffin-embedded tissues among invasive breast cancer cases identified from the Nurses' Health Study (NHS). Tissue microarrays were constructed from archived tissue blocks of women diagnosed with breast cancer in the NHS (1976–1997). Invasive non-metastatic breast cancer tumors ( $n = 1,945$ ) were classified into 1 of 5 molecular phenotypes based on immunohistochemistry assays for estrogen receptor (ER), progesterone receptor (PR), HER2, cytokeratin (CK) 5/6, epidermal growth factor receptor

(EGFR) and grade. Survival outcomes were estimated using the Kaplan–Meier product limit method. Cox-proportional hazards models were fitted to determine the association of molecular phenotype with survival outcomes after adjusting for covariates. 1,279 (65.8%) tumors were classified as luminal A, 279 (14.3%) as luminal B, 95 (4.9%) as HER2 type, 203 (10.4%) as basal-like and 89 (4.6%) tumors were unclassified. The 5-year breast cancer-specific survival estimates for women with luminal A, luminal B, HER2-type, basal-like and unclassified tumors were 96, 88, 81, 89 and 85%, respectively. In the multivariable model, compared to cases with luminal A tumors, cases with luminal B (HR 1.90, 95% CI 1.33–2.71), HER2-type (HR 1.36, 95% CI 0.87–2.12), basal-like (HR 1.58, 95% CI 1.05–2.39) and unclassified (HR 1.38, 95% CI 0.87–2.20) tumors had higher hazard of breast cancer death. Similar trends were observed for both overall and recurrence-free survival. In conclusion, compared to women who have luminal A tumors those with luminal B, HER2-type, basal-like and unclassified tumors had a worse prognosis, when tumor subtype was defined by immunohistochemistry. This method may provide a cost-effective means of determining prognosis in the clinical setting.

## Keywords

Breast cancer; Molecular phenotypes; Survival

## Introduction

Breast cancer is known to be a heterogeneous disease with microarray profiling studies having identified several biologically distinct subtypes of breast tumors, each associated with different clinical outcomes [1–4]. The basal-like and HER-2 over-expressing subtypes are associated with poorer prognosis relative to women with luminal tumors. Among the hormone receptor positive subtypes, women with luminal B tumors have a significantly worse prognosis compared to women with luminal A tumors. These distinct subtypes have been shown to be conserved across diverse patient series and array platforms [5,6]. Using a single data set of 295 samples, Fan et al. [7] was also able to demonstrate significant agreement across different gene expression-based predictors implying that these gene sets probably track a common set of biological phenotypes.

Issues such as cost, complexity and technical expertise have limited the use of gene expression profiling as a routine diagnostic tool in the hospital setting. The next best alternative is to use antibodies that work on formalin-fixed, paraffin-embedded tissue. Immunohistochemistry panels have been proposed to classify breast tumors into distinct subtypes as identified by gene expression profiling studies [8–14]. These panels primarily use antibodies against estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), cytokeratin 5/6 (CK5/6) and epidermal growth factor receptor (EGFR). These panels also allow the use of archived tissue samples that have associated long-term clinical follow-up data such as that available among women enrolled in the Nurses' Health Study (NHS). In this study, we used archived breast tissue specimens of women with invasive breast cancer enrolled into the NHS and used immunohistochemistry panels to classify tumors according to breast tumor subtype and examined the relationship between these subtypes and long-term survival outcomes.

## Materials and methods

### Study population and breast cancer case identification

The NHS cohort was established in 1976 when 121,700 female registered nurses aged between 30 and 55 years from across the United States were recruited upon answering a mailed questionnaire that was aimed at determining risk factors for cancer and

cardiovascular disease. Follow-up questionnaires have since been sent out every 2 years to the study participants to update exposure information and determine development of non-fatal malignant and non-malignant disease with the follow-up rate being consistently high approaching over 90%.

Among study participants reporting an incident diagnosis of breast cancer, written permission was obtained to review their medical records to both confirm diagnosis and record information on tumor characteristics including whether the cancer was in situ or invasive, histological subtype, tumor size and the presence or absence of metastases. Using this method, 99% of self reported breast cancer cases was confirmed. To identify cases of breast cancer among participants who had died, death certificates were obtained to ascertain the cause of death and thus identify and confirm those attributed to breast cancer.

This study was approved by the Human Subjects Committee of the Brigham and Women's Hospital in Boston, MA. Completion of the self-administered questionnaire was considered to imply informed consent.

### **Tissue block collection and molecular phenotype classification**

Details of tissue block collection, construction of tissue microarrays (TMA), immunohistochemical analysis and classification into various breast cancer phenotypes have been described in detail elsewhere [15]. In summary, in 1993, the NHS began collecting archived formalin-fixed paraffin-embedded breast cancer blocks from study participants with confirmed primary incident breast cancers over 20 years of follow-up (1976–1996). Of the 5,610 women with breast cancer that were eligible for block collection, pathology samples were available and obtained for 3,752 participants. Of these, 23 TMA blocks were constructed from 3,093 cancers and positive lymph nodes from 2,897 participants. Participants who reported having breast cancer at the time of study entry were excluded.

TMAs were constructed at the Dana Farber Harvard Cancer Center Tissue Microarray Core Facility in Boston. Cores of 0.6 mm were obtained from the tissue blocks containing breast tumors and inserted into recipient TMA blocks. Subsequently, 5- $\mu$ m paraffin sections were cut from the TMA block that underwent immunohistochemical staining for ER, PR, HER2, CK5/6 and EGFR. Nuclei of tumor cells exhibiting staining for ER or PR either at a low level (1–10% of tumor cell nuclei staining) or high level (>10% tumor cell nuclei staining) was considered to be positive for ER and PR, respectively. ER and PR negative tumors were defined as those that exhibited complete absence of tumor cell staining. Tumors were considered as HER2 positive when more than 10% of tumor cells showed moderate or strong membrane staining (2+ and 3+). Tumor exhibiting 0 or 1+ staining for HER2 protein over-expression was considered to be HER2 negative. To classify tumors as being CK-positive or EGFR-positive, any degree of cytoplasmic and/or membranous staining, even if focal, was required.

Using the results of the immunohistochemical analyses, we classified tumors into five subtypes. Cases that were ER-positive and/or PR-positive, HER2-negative and grades 1 and 2 were classified as luminal A cancers; cases that were either (a) ER-positive and/or PR-positive and HER2-positive or (b) ER-positive and/or PR-positive and HER2-negative and high grade were classified as luminal B cancers. Tumors that exhibited negative staining for both ER and PR, positive staining for HER2 protein were classified as HER2 type. Basal-like tumors were defined as those that exhibited no staining for ER, PR and HER2 and positive staining for CK5/6 and/or EGFR. If the tumor exhibited no staining for all five markers the tumor was categorized as “unclassified”.

## Inclusion and exclusion criteria of study participants

Inclusion criteria for this study were women enrolled in the NHS who had stages I–III invasive breast cancer diagnosed between 1976 and 1996 that was confirmed by medical record review and whose breast tumor blocks were available for TMA construction and subsequent immunohistochemical analysis. We excluded participants with positive lymph nodes only ( $n = 25$ ), rare tumor types including malignant phyllodes tumors, neuroendocrine carcinoma and angiosarcoma ( $n = 10$ ), breast cancer cases with in situ ( $n = 401$ ), stage IV disease ( $n = 62$ ), missing information on molecular phenotype ( $n = 225$ ), metastases at diagnosis or stage III but lacking a complete metastatic work-up ( $n = 175$ ), and women with impossible date of recurrence which was estimated prior to date of diagnosis ( $n = 54$ ).

## Calculation of mortality and breast cancer recurrences

Women were followed until death or December 2007 whichever ever came first. Deaths were determined by reporting from family members and/or postal authorities or searching the National Death Index. Using the National Death Index is a known and reliable method for women with breast cancer [16,17]. Almost 98% of the deaths among women in the NHS have been determined in this fashion [18]. Date of death was ascertained from death certificates. Cause of death was determined from death certificates and/or review of medical records.

In this analysis, we assumed breast cancer cases recurred if they subsequently reported a second cancer in the liver, bone or brain as these are the most common sites of breast cancer recurrences. In cases where women with breast cancer reported the development of a subsequent lung cancer, medical records were reviewed to distinguish breast cancer metastases to the lung from a primary lung cancer. Among women whose cause of death was reported to be due to breast cancer, recurrence was assumed to have occurred 2 years prior to the date of death. This assumption is based on the fact that the average survival of women with stage IV disease is approximately 2 years [19].

## Statistical analysis

Information on covariates was obtained from questionnaires that are completed by the participants biennially. For the purposes of this study covariate information at the time of diagnosis was obtained from the questionnaire before the report of breast cancer diagnosis. The covariates were compared across molecular phenotype groups with the Chi-square test or Wilcoxon's rank sum test, as appropriate. Median follow up was measured as the median observation time among all cases.

Three survival end points were defined for this study with the follow-up cut off being December 2007. Overall survival (OS) was calculated from the date of diagnosis to the date of death from any cause or the follow-up cut off whichever came first. Breast cancer-specific survival (BCS) was calculated from the date of diagnosis to the date of death from breast cancer or the follow-up cut off. Recurrence-free survival (RFS) was calculated from the date of diagnosis to the date of first metastatic recurrence or the follow-up cut off. For the estimation of BCS and RFS, deaths from any other causes were censored. The Kaplan–Meier product limit method was used to estimate the three survival end points and was compared across groups using log rank statistic. Cox-proportional hazards models were then fit to determine the association of molecular phenotype with each survival end point after adjusting for lifestyle and tumor characteristics.

Covariate information on the study population was obtained from biennial questionnaires. The following covariate data were obtained from the questionnaire preceding the report of breast cancer diagnosis: age, body mass index (BMI), oral contraceptive use, age at first

birth, parity, postmenopausal hormone use, alcohol intake and smoking status. Information on breast tumor characteristics and treatments was extracted from the medical record and supplemental questionnaire including year of diagnosis, stage, radiation, and chemotherapy and hormonal treatment. Information on histological grade was obtained from centralized pathology review by a single pathologist (YF). Covariates considered in the multivariate model were based on both statistical significance and clinical significance. Variables included in the final model were age at diagnosis (continuous), year of diagnosis (continuous), body mass index at diagnosis (continuous), grade (grades 1, 2 and 3), stage of disease (stages I, II and III), radiation treatment (yes, no, missing), and chemotherapy and hormonal treatment (no/no, yes/no, no/yes, yes/yes, missing). The proportional hazards assumption of the final model was assessed visually with plots of the model residuals and with interaction terms of the covariates with time. All statistical tests were two-sided and *P* values less than 0.05 were considered statistically significant. Analyses were performed using the SAS 9.1.

## Results

The final analyses included a total of 1,945 women with non-metastatic breast cancer at the time of diagnosis. Table 1 summarizes characteristics of the women enrolled in the NHS and their tumors according to the breast cancer molecular phenotypes. Mean age at breast cancer diagnosis ranged from 55 to 58 years across the five molecular phenotypes with women diagnosed with luminal A tumors tending to be older at diagnosis compared to the other molecular phenotypes. 1,279 (65.8%) tumors were classified as luminal A, 279 (14.3%) were classified as luminal B, 95 (4.9%) were of HER2 type, 203 (10.4%) were classified as basal-like and 89 (4.6%) tumors were unclassified. In general, women with luminal A tumors had tumors that were smaller, lower grade and stage, and with less nodal involvement compared with the other molecular phenotypes.

### Survival estimates

Median follow-up among all women was 15 years. At the time of the analysis (follow-up cut off December 2007), 728 (37.4%) women had died of any cause, 395 (20.3%) had died of a breast cancer related event, and 417 (21.4%) experienced a recurrence. Table 2 summarizes the 5- and 10-year survival estimates for the five molecular phenotypes. Five-year BCS for women with luminal A, luminal B, HER2 type, basal-like and unclassified tumors was 96, 88, 81, 89 and 85%, respectively ( $P < 0.0001$ ) (Fig. 1). Similar trends were observed for both OS and RFS.

### Multivariable model

Table 3 summarizes the age adjusted and multivariable adjusted hazard ratios for OS, BCS and RFS according to the five molecular phenotypes. The multivariable model was adjusted for age and year of diagnosis, BMI at diagnosis, stage of disease, nodal status, tumor size and grade of disease. Compared to cases with luminal A tumors cases with luminal B (HR 1.90, 95% CI 1.33–2.71), HER2 type (HR 1.36, 95% CI 0.87–2.12), basal-like (HR 1.58, 95% CI 1.05–2.39) and unclassified (HR 1.38, 95% CI 0.87–2.20) tumors had higher hazard of breast cancer death. Compared to cases with luminal A tumors cases with luminal B (HR 1.77, 95% CI 1.25–2.52), HER2-type (HR 1.21, 95% CI 0.78–1.89), basal-like (HR 1.37, 95% CI 0.91–2.06) and unclassified (HR 1.51, 95% CI 0.97–2.35) tumors had higher hazard of breast cancer recurrence.

## Discussion

Using a large prospective cohort study, we examined the prognosis associated with various invasive breast tumor phenotypes, as defined by a panel of immunohistochemical markers. The results of this large study, with relatively long follow-up, reveal that women with luminal A tumors have better OS, BCS and RFS than women with other molecular phenotypes after adjustment for a number of clinically relevant potential confounders.

The results of this study are similar to those reported in earlier studies that used either gene expression profiling [1,2,4] or panels of immunohistochemical markers [8–14]. In one of the earliest studies by Sorlie et al. [1], gene expression profiling was used to classify 48 invasive breast tumor samples into various subtypes. At a median follow-up of 66 months the authors were able to show in a univariate analysis that the subgroup of women with HER2-type and basal-like tumors had worse prognosis compared to women with luminal breast tumors. Using invasive tumors samples from a cohort of 4,046 women, Cheang et al. [12] classified these tumors into various phenotypic subtypes using a panel of five immunohistochemical markers and found that women with luminal tumors had better survival outcomes compared to those whose tumors were of HER2-type or basal-like. Similarly, the current study with the longest median follow-up to date also found that HER2-type and basal-like tumors as defined by a panel of immunohistochemical markers had worse survival relative to the luminal A type.

Most studies evaluating prognostic outcomes related to breast tumor subtype, including the one presented here, were conducted in the pre-trastuzumab era. Trastuzumab efficacy trials have demonstrated significant survival benefit in both the metastatic [19] and adjuvant setting [20–23]. Recently, Dawood et al. [24] presented results of a large retrospective study of approximately 2,000 women with metastatic breast cancer demonstrating that women with HER2 over-expressing breast cancers had a 44% reduction in the risk of death (HR 0.56, 95% CI 0.45–0.69,  $P < 0.0001$ ) compared to women with HER2 non-expressing tumors. Thus, it is unclear if the results relating to HER2-expressing tumors are applicable to women being treated in the trastuzumab era.

Sorlie et al. [1] demonstrated that gene expression profiling of breast tumors resulted in a fifth subtype that the authors classified as “normal breast-like”, which constituted approximately 6% of breast tumors and was associated with good prognosis. These tumors were found to have high expression of many genes known to be expressed in adipose tissue and also demonstrated strong expression of basal epithelial genes as well as low expression of luminal epithelial genes. In our cohort approximately 5% of women had tumors that were determined to be “unclassified” due to the absence of staining of their tumors for ER, PR, HER2, EGFR and CK 5/6. 5 and 10-year BCS among this subgroup of women was 85 and 82%, respectively. After adjusting for a number of clinically relevant potential confounders we noted that compared to women with luminal A tumors those with basal-like tumors had a 58% increased risk of breast cancer-specific death while those with unclassified tumors had a 38% increased risk of breast cancer-specific death. The current study is consistent with previous studies which have examined these breast cancer subtypes [12,14]. In the Carolina Breast Cancer Study ( $n = 496$  incident cases of invasive breast cancer), Carey et al. [14] found that women whose tumors did not express any of the immunohistochemical markers used to classify breast tumors (ER, PR, HER2, EGFR, and CK5/6) had 77% breast cancer survival over 11.2 years. Similarly, Cheang et al. [12] recently reported a 10-year BCS of 72% among women with “unclassified” tumors as determined by immunohistochemistry.

When interpreting the results of our study it is important that we do so within the confines of its strengths and limitation. Detailed treatment information was unavailable in our study;

however, we were able to adjust for general treatment categories in the multivariable models. The importance of treatment is highlighted by the recent 15-year update from Early Breast Cancer Trialists' Collaborative group that showed that 6 months of adjuvant anthracycline chemotherapy reduced the annual breast cancer death rate by 38% for women younger than 50 years and by 20% for those women aged 50–69 years [25]. Although it is possible that there could be residual confounding by treatment, we attempted to account for changes in treatment over time by including year of diagnosis in our multivariable model. Women contributing to this analysis only included those whose tumor samples were available for inclusion in the TMA. After taking into account age and year of diagnosis the women included in this analysis were similar to the women whose tumor samples were unavailable for TMA [15], suggesting that these women are missing at random.

## Conclusion

The results of our study add to the growing literature that classifies invasive breast tumors into various subtypes similar to that identified by gene expression profiling. The strength of this study lies in the fact that the information derived was acquired prospectively among a large group of health professionals diagnosed and treated in the community setting with long follow-up. The cost and technical expertise required to classify breast tumors will in all probability preclude its use in the community setting and as such classification based on immunohistochemistry presents an attractive option. With the movement towards personalized treatment programs classification of breast tumors will be even more important. However, limitations still exist. Future studies will need to focus improved classification of tumors using immunohistochemical markers.

## Abbreviations

<b>ER</b>	Estrogen receptor
<b>PR</b>	Progesterone receptor
<b>HER2</b>	Human epidermal growth factor receptor 2
<b>CK5/6</b>	Cytokeratin 5/6
<b>EGFR</b>	Epidermal growth factor receptor
<b>NHS</b>	Nurses' Health Study
<b>TMA</b>	Tissue microarray
<b>OS</b>	Overall survival
<b>BCS</b>	Breast cancer specific survival
<b>RFS</b>	Recurrence-free survival
<b>BMI</b>	Body mass index
<b>HR</b>	Hazard ratio
<b>95% CI</b>	95 percent confidence interval

## Acknowledgments

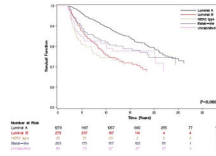
We thank Dr. Yineng Fu for central pathology review of all the breast cancer cases included in this study. We are grateful to the participants of the Nurses' Health Study for their outstanding dedication and commitment to the study. This study was supported by GlaxoSmithKline (WE234 (EPI40307)), Public Health Service Grants CA087969, and SPORE in Breast Cancer CA089393, from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services and Breast Cancer Research Fund. Dr. Graham Colditz was supported in part by an American Cancer Society Cissy Hornung Clinical Research Professorship.

## References

1. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001;98:10869–10874. [PubMed: 11553815]
2. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2001;100:8418–8423. [PubMed: 12829800]
3. Van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature (Lond.)* 2002;415:530–536. [PubMed: 11823860]
4. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population based study. *Proc Natl Acad Sci USA* 2003;100:10393–10398. [PubMed: 12917485]
5. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96. [PubMed: 16643655]
6. Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* 2006;7:127. [PubMed: 16729877]
7. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–569. [PubMed: 16899776]
8. Nielson TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast cancer. *Clin Cancer Res* 2004;10:5367–5374. [PubMed: 15328174]
9. Abd El-Rehim DM, Ball G, Pinder SE, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterized series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int J Cancer* 2005;116:340–350. [PubMed: 15818618]
10. Abd El-Rehim DM, Pinder SE, Paish CE, et al. Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* 2004;203:661–671. [PubMed: 15141381]
11. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002;161:1991–1996. [PubMed: 12466114]
12. Cheang MCU, Voduc D, Bajdik C, et al. Basal-like breast defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res* 2008;14:1368–1376. [PubMed: 18316557]
13. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;13:2329–2334. [PubMed: 17438091]
14. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492–2502. [PubMed: 16757721]
15. Tamimi RM, Baer HJ, Marotti J, Galan M, Galaburda L, Fu Y, et al. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Res* 2008;100:218–221.
16. Sathiakumar N, Delzell E, Abdalla O. Using the National Death Index to obtain underlying cause of death codes. *J Occup Environ Med* 1998;40:808–813. [PubMed: 9777565]
17. Rutqvist LE. Validity of certified causes of death in breast carcinoma patients. *Acta Radiol Oncol* 1985;24:385–390. [PubMed: 3002134]
18. Stampfer MJ, Willett WC, Speizer FE, et al. Test of the National Death Index. *Am J Epidemiol* 1984;119:837–839. [PubMed: 6720679]
19. Harris, J.; Lippman, M.; Morrow, M.; Osborne, C. *Diseases of the breast*. 3rd edn.. Philadelphia: Lippincott; 2004.



20. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–792. [PubMed: 11248153]
21. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;353:1673–1684. [PubMed: 16236738]
22. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;353:1659–1672. [PubMed: 16236737]
23. Joensuu H, Kellokumpu-Lehtinen PL, Bono P, et al. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N Engl J Med* 2006;354:809–820. [PubMed: 16495393]
24. Dawood S, Kristine B, Hortobagyi GN, Giordano SH. Prognosis of women with stage IV breast cancer by HER2 status and trastuzumab treatment: An institutional based review. *J Clin Oncol* 2008;26 suppl:1018.
25. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomized trials. *Lancet* 2005;365:1687–1717. [PubMed: 15894097]



**Fig. 1.** Kaplan–Meier plots for breast cancer-specific survival by molecular phenotype among invasive breast cancer identified from the Nurses' Health Study ( $P < 0.0001$ )

**Table 1**  
Descriptive characteristics of breast cancer cases and tumors according to molecular phenotype (N = 1,945)

	Luminal A	Luminal B	HER2-type	Basal-like	Unclassified	P value
Number	1,279 (65.8%)	279 (14.3%)	95 (4.9%)	203 (10.4%)	89 (4.6%)	–
Age at diagnosis (years)	58.1	56.7	56.8	55.4	54.9	<0.0001
Chemotherapy						
Yes	299 (30.1%)	97 (47.8%)	41 (60.3%)	99 (63.1%)	28 (48.3%)	
No	694 (69.9%)	106 (52.2%)	27 (39.7%)	58 (36.9%)	30 (51.7%)	<0.0001
Missing	286	76	27	46	31	
Tamoxifen						
Yes	715 (71.9%)	150 (74.3%)	32 (49.2%)	43 (27.7%)	25 (42.4%)	
No	280 (28.1%)	52 (25.7%)	33 (50.8%)	112 (72.3%)	34 (57.6%)	<0.0001
Missing	284	77	30	48	30	
Radiation therapy						
Yes	431 (43.1%)	80 (39.8%)	23 (34.3%)	72 (46.5%)	25 (43.1%)	
No	570 (56.9%)	121 (60.2%)	44 (65.7%)	83 (53.5%)	33 (56.9%)	0.4702
Missing	278	78	28	48	31	
Nodal involvement						
None	969 (75.8)	188 (67.4)	56 (59.0)	140 (69.0)	66 (74.2)	
1–3	252 (19.7)	68 (24.4)	29 (30.5)	52 (25.6)	15 (16.9)	
4–9	31 (2.4)	14 (5.0)	6 (6.3)	8 (3.9)	5 (5.6)	
≥10	27 (2.1)	9 (3.2)	4 (4.2)	3 (1.5)	3 (3.4)	0.005
Tumor size (cm)						
<2	934(73.0%)	161 (57.7%)	48 (50.5%)	114 (56.2%)	57 (64.0%)	
≥2	345 (27.0%)	118 (42.3%)	47 (49.5%)	89 (43.8%)	32 (36.0%)	<0.0001
Stage of disease						
I	766 (59.9%)	127 (45.5%)	36 (37.9%)	91 (44.8%)	49 (55.1%)	
II	429 (33.5%)	117 (41.9%)	44 (46.3%)	97 (47.8%)	31 (34.8%)	
III	84 (6.6%)	35 (12.5%)	15 (15.8%)	15 (7.4%)	9 (10.1%)	<0.0001
Grade of tumor						
I	385 (30.1%)	7 (2.5%)	2 (2.2%)	6 (3.0%)	18 (20.5%)	
II	894 (69.9%)	52 (18.8%)	43 (47.8%)	50 (24.9%)	31 (35.2%)	

	Luminal A	Luminal B	HER2-type	Basal-like	Unclassified	P value
III	0	217 (78.6%)	45 (50.0%)	145 (72.1%)	39 (44.3%)	<0.0001

Table 2

5- and 10-year survival estimates by molecular phenotype of the tumor

	Luminal A	Luminal B	HER2-type	Basal-like	Unclassified	P value
Overall survival						
5-year (%) (95% CI)	94 (92–95%)	85 (81–89%)	80 (72–88%)	86 (82–91%)	82 (74–90%)	
10-year (%) (95% CI)	82 (80–85%)	71 (65–76%)	73 (64–82%)	77 (72–83%)	75 (66–84%)	P = 0.0096
Breast cancer-specific survival						
5-year (%) (95% CI)	96 (95–97%)	88 (84–92%)	81 (73–89%)	89 (85–93%)	85 (77–93%)	
10-year (%) (95% CI)	90 (88–91%)	76 (71–81%)	74 (66–83%)	82 (77–88%)	81 (73–90%)	P <0.0001
Recurrence-free survival						
5-year (%) (95% CI)	93 (92–95%)	82 (78–87%)	78 (69–86%)	86 (81–90%)	81 (72–89%)	
10-year (%) (95% CI)	87 (85–89%)	74 (69–79%)	74 (65–82%)	80 (74–86%)	73 (64–83%)	P <0.0001

Table 3

Age-adjusted and multivariable hazard rates of death, breast cancer death and recurrence according to molecular phenotype

	Luminal A	Luminal B	HER2-type	Basal-like	Unclassified
<i>N</i> (total = 1,945)	1,279 (65.8%)	279 (14.3%)	95 (4.9%)	203 (10.4%)	89 (4.6%)
Total deaths = 728					
Age adjusted HR (95% CI)	1.00 (-)	1.47 (1.21–1.79)	1.34 (0.97–1.84)	1.14 (0.89–1.46)	1.13 (0.80–1.62)
Multivariate HR (95% CI)	1.00 (-)	1.45 (1.09–1.93)	1.12 (0.78–1.60)	1.23 (0.89–1.69)	1.18 (0.81–1.70)
Breast cancer deaths = 395					
Age adjusted HR (95% CI)	1.00 (-)	1.79 (1.39–2.32)	1.76 (1.18–2.63)	1.29 (0.94–1.78)	1.38 (0.88–2.17)
Multivariate HR (95% CI)	1.00 (-)	1.90 (1.33–2.71)	1.36 (0.87–2.12)	1.58 (1.05–2.39)	1.38 (0.87–2.20)
Recurrences = 417					
Age adjusted HR (95% CI)	1.00 (-)	1.75 (1.36–2.24)	1.66 (1.12–2.47)	1.21 (0.88–1.66)	1.54 (1.01–2.34)
Multivariate HR (95% CI)	1.00 (-)	1.77 (1.25–2.52)	1.21 (0.78–1.89)	1.37 (0.91–2.06)	1.51 (0.97–2.35)

Multivariable models adjusted for age at diagnosis, diagnosis period, body mass index, disease stage (I, II, III), grade (I, II, III), radiation treatment (yes, no, missing), chemotherapy and hormonal treatment (no/no, yes/no, no/yes, yes/yes, missing)